

AMENDMENTS TO THE CLAIMS:

Claims 1-15 and 17-28 are pending in the subject application. Each of claims 1-4, 7, 14, 19, 20 and 22-28 has been amended herein. This Listing of Claims will replace all prior versions and listings of claims in the application:

Listing of Claims:

1. (Currently Amended) A system that facilitates incremental web crawls comprising the following components stored in computer memory and executable by a processor:

an indexer that places items with similar properties into respective chunks, wherein the properties are shared by all the items within a respective chunk, wherein the items are the results returned by a web crawl, and wherein the respective chunks include at least one rank chunk and at least one webmap chunk; and[[,]]

a chunk map that stores at least some of the properties associated with ~~the~~ each respective chunk, ~~the stored properties are shared by all the items in the respective chunk,~~ wherein the properties include ~~are at least one of~~ average time between change and average importance of documents in ~~the~~ each respective chunk, wherein the chunk map is employed to facilitate an incremental web re-crawl, and wherein the properties of each respective chunk stored in the chunk map are utilized to determine a re-crawl of all the items in that respective chunk.

2. (Currently Amended) The system of claim 1, wherein the items comprise ~~comprising~~ information associated with a Uniform Resource Locator.

3. (Currently Amended) The system of claim 1, wherein the items comprise ~~comprising~~ at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file.

4. (Currently Amended) The system of claim 1, wherein the items ~~receives~~ are received from a crawler, and wherein the crawler is responsible for a specific set of Uniform Resource Locators.

5. (Original) The system of claim 1, further comprising a master control process that can modify the chunk map to facilitate load balancing amongst a plurality of crawlers.

6. (Original) The system of claim 1, further comprising a master control process that serves as an interface between a crawler and a re-crawl controller.

7. (Currently Amended) The system of claim 6, wherein the master control process maintains a known chunks table that stores information for components of ~~[[a]]~~ the system.

8. (Original) The system of claim 6, wherein the master control process exposes an interface for communication with a component of the system.

9. (Original) The system of claim 8, wherein the interface returns a list of chunks the component should have and where to get the chunks.

10. (Original) The system of claim 8, wherein the interface returns a list of the chunks that should be actively served by the component.

11. (Original) The system of claim 8, wherein the interface returns a range of chunk identifiers to use in building a new chunk by the component.

12. (Original) The system of claim 8, wherein the interface causes an old chunk to be retired by the system.

13. (Original) The system of claim 6, wherein the master control process facilitates movement of chunks from one component to another component.

14. (Currently Amended) The system of claim 13, wherein movement of chunks is based, at least in part, upon at least one of rebalancing index servers after one goes down, re-crawling pages previously crawled, and[[,]] restoring a state of a the crawler after it has crashed.

15. (Original) The system of claim 1, further comprising a re-crawl component that employs the chunk map to determine which chunks, if any, to re-crawl at a particular time.

16. (Cancelled)

17. (Original) The system of claim 1, further comprising an index chunk that stores information associated with an index of at least some of the items.

18. (Original) The system of claim 1, ~~further comprising a~~ wherein the at least one rank chunk that stores a static rank associated with an index chunk.

19. (Currently Amended) A method of performing document re-crawl comprising:

selecting a first chunk from a group that includes at least one index chunk, at least one rank chunk, at least one content chunk, at least one re-crawl chunk, and at least one webmap chunk, wherein a chunk map that stores properties associated with the first chunk is employed to determine the first chunk based on the stored properties, and wherein the stored properties include average time between change and average importance and are shared by all items in the first chunk;

parsing the [[a]]first chunk for ~~uniform resource locators~~Uniform Resource Locators, wherein a chunk map that stores properties associated with the respective chunk is employed to determine the first chunk, wherein the stored properties are shared by all the items in the respective chunk;

re-crawling the ~~uniform resource locators~~Uniform Resource Locators;

receiving one or more documents as a result of re-crawling the uniform resource locators;

storing a document from the one or more documents in an appropriate chunk when it is determined that the document belongs to the appropriate chunk;
and,

forming a second chunk separate from the first chunk[[,]] based, at least in part, upon the re-crawled uniform resource locators when it is determined that the document does not belong to any chunk from the group, and storing the document in the second chunk.

20. (Currently Amended) The method of claim 19, further comprising at least one of the following acts:

determining whether any chunks are to be retired;
moving the first chunk; and[[,]]
destroying the first chunk.

21. (Original) One or more computer readable media having stored thereon computer executable instructions for carrying out the method of claim 19.

22. (Currently Amended) A method of performing document re-crawl comprising:

parsing a first chunk for ~~uniform resource locators~~ Uniform Resource Locators, wherein the ~~uniform resource locators~~ Uniform Resource Locators are stored as a result of one or more web crawls;

accessing a chunk map comprising properties associated with one or more chunks that include the ~~respective first~~ first chunk, wherein the stored properties are shared by all the items in the ~~respective first~~ first chunk, and wherein the properties include ~~are at least one of~~ average time between change and ~~or~~ average importance of documents in the ~~respective first~~ first chunk; and,

periodically determining, based on the properties of each of the one or more chunks ~~chunk~~ in the chunk map, whether to re-crawl the items in the first chunk of data.

23. (Currently Amended) The method of claim 22, wherein the ~~period~~ periodic determination is ~~being~~ based, at least in part, upon[[,]] at least one of average time between change and average importance of documents comprising a particular chunk.

24. (Currently Amended) A computer-readable medium encoded with a data structure transmitted between two or more computer components that facilitates document re-crawl, the data structure comprising:

a chunk header that includes metadata associated with the data structure, wherein the metadata is shared by all the items in a chunk that comprises document files associated with one or more uniform resource locators, wherein properties of all the document files include average time between change and importance of a document;

an offset section that provides offset information associated with the document files; ~~and,~~

wherein the document files ~~that~~ include content found on the Internet, wherein an ~~the~~ average of ~~the~~ at least one of the properties of all the document files determines if the document should be re-crawled,

and wherein data in the data structure is compressed.

25. (Currently Amended) The data structure of claim 24, wherein at least one of the document files comprises ~~comprising~~ at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file.

26. (Currently Amended) A system that facilitates ~~increment~~ incremental web crawls comprising the following components stored in computer memory and executable by a processor:

means for placing items with similar properties into respective chunks, wherein the properties are shared by all the items in a respective chunk, and wherein the items are results returned by a web crawl; ~~and,~~

means for storing at least some of the properties associated with ~~the~~ each respective chunk, ~~wherein the stored properties are shared by all the items in the respective chunk, and~~ wherein the properties include ~~are at least one of~~ average time between change and ~~or~~ average importance of documents in the respective chunk[[,]]; and

means for employing the stored properties of each respective chunk to facilitate an incremental web re-crawl.

27. (Currently Amended) The system of claim 26, wherein the items comprise ~~comprising~~ information associated with a Uniform Resource Locator.

28. (Currently Amended) The system of claim 26, wherein the items comprise ~~comprising~~ at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file.